

(19) 日本国特許庁 (J P)

(12) 公表特許公報 (A)

(11) 特許出願公表番号

特表平10-508389

(43) 公表日 平成10年(1998) 8月18日

(51) Int.Cl.<sup>6</sup>

G 1 0 L 3/00

識別記号

5 1 3

F I

G 1 0 L 3/00

5 1 3 B

5 1 3 A

7/08

7/08

A

9/10

3 0 1

9/10

3 0 1 C

審査請求 未請求 予備審査請求 未請求(全 25 頁)

(21) 出願番号 特願平8-504873  
(86) (22) 出願日 平成6年(1994) 7月18日  
(85) 翻訳文提出日 平成8年(1996) 3月14日  
(86) 国際出願番号 PCT/J P 94/0 1 1 8 1  
(87) 国際公開番号 WO 9 6/0 2 9 1 1  
(87) 国際公開日 平成8年(1996) 2月1日  
(81) 指定国 CN, J P, KR, US

(71) 出願人 松下電器産業株式会社  
大阪府門真市大字門真1006番地  
(71) 出願人 パナソニック テクノロジズ インク  
米国 08540 ニュージャージー プリンストン サードフロア2 リサーチウェイ  
(72) 発明者 リーヴィス ベンジャミン カー  
奈良県大和高田市築山492-23  
(74) 代理人 弁理士 松田 正道

(54) 【発明の名称】 音声検出装置

(57) 【要約】

装置は、入力信号内に含まれる音声の開始及び終了部を該信号内の平滑化周波数帯域制限エネルギーの分散及び平滑化周波数帯域制限エネルギーの履歴に基づいて検出する。分散の使用は、前記信号を有する絶対的信号対雑音比とは比較的無関係である検出を可能にし、音楽、モータ雑音及び他の音声のような背景雑音のような広範囲のいろいろな背景内での正確な検出を可能にする。装置は、高速の特別目的のデジタル信号プロセッサ集積回路に加えて既製のハードウェアを使用して容易に実施され得る。

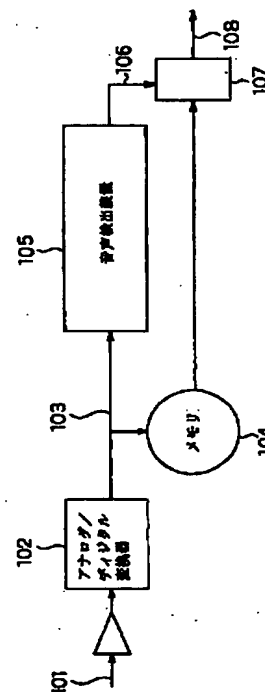


図 1

**【特許請求の範囲】**

1. 入力信号における音声を検出する装置において、  
前記信号内の平滑化周波数帯域制限エネルギーを表す値を決定する手段と、  
  
平滑化周波数帯域制限エネルギーの分散を決定する手段と、  
前記平滑化周波数帯域制限エネルギーの分散及び前記平滑化周波数帯域制限エネルギーの以前の履歴に基づいて前記信号内の音声の開始及び終了点を決定する手段とを備えていることを特徴とする入力信号における音声を検出する装置。  
  
2. 前記平滑化周波数帯域制限エネルギーを表す値を決定する前記手段が、  
前記信号に関連した周波数を決定する手段と、  
予め選択された範囲内の周波数を有する前記信号の一部を選択する手段と  
  
前記信号の選択された一部内の全エネルギーを表す値を決定する手段であって、前記全エネルギーを表す値が前記周波数帯域制限エネルギーであることと  
  
前記周波数帯域制限エネルギーを平滑化する手段であって、前記値が前記平滑化周波数帯域制限エネルギーであることと、  
を備えていることを特徴とする請求の範囲1の装置。  
  
3. 前記平滑化周波数帯域制限エネルギーを表す値を決定する前記手段が、  
ハミングウィンドウフィルタを信号の一部に適用し、フィルタリングされ信号を生成する手段と、  
  
フーリエ変換を前記フィルタリングされた信号に適用し、変換された信号を生成する手段と、  
  
前記変換された信号を合計し、前記信号の一部における前記全エネルギーを表す値を生成する手段であって、前記信号のエネルギーを表す値が前記周波数帯域制限エネルギーであることと、  
  
フィルタを前記周波数帯域制限エネルギーを適用する手段であって、結果が前記平滑化周波数帯域制限エネルギーであることと、

を備えていることを特徴とする請求の範囲1の装置。

4. 装置が、

前記音声信号を受け取る手段と、

m秒の連続期間に及ぶ信号の一部を記憶する手段と、

新しい信号が受け取られたとき前記信号の記憶された一部を更新する手段とを含んでいることを特徴とする請求の範囲1の装置。

5. mが0と10秒の間であることを特徴とする請求の範囲4の装置。

6. 前記信号の一部を記憶する前記手段がシフトレジスタを備えていることを特徴とする請求の範囲4の装置。

7. 前記平滑化周波数帯域制限エネルギーの分散を決定する前記手段が、

前記平滑化周波数帯域制限エネルギーを表す複数の値を記憶する手段であって、前記値が時間の関数として記憶されていることと、

$V$ が $V = g(A, B)$ が与えられるところの分散 $V$ を計算する手段であって、ここで、 $BLE(f)$ が平滑化周波数帯域制限エネルギーの複数の値、 $nv$ が値の数、 $f = nv, \dots, 3, 2, 1$ 、 $BLE(1)$ が最古の $BLE$ 値であることと、

を備えていることを特徴とする請求の範囲1の装置。

8. 前記平滑化周波数帯域制限エネルギーの分散を決定する前記手段が、

新しい $BLE(nv)$ の値が受け取られるとき、 $V = g(A', B')$ を計算する手段であって、

ここで、

$$A' = A + [BLE(nv) \times BLE(nv)] - [BLE(0) \times BLE(0)]$$

$$B' = B + BLE(nv) - BLE(0)$$

ここで、

$A'$  は $A$ に対する更新値であり、

$B'$  は $B$ に対する更新値であり、

及び

$BLE(nv)$ が最新の平滑化周波数帯域制限エネルギーであり、かつ

$BLE(0)$ が最古の平滑化周波数帯域制限エネルギーであることと、

をさらに備えていることを特徴とする請求の範囲7の装置。

9. 前記平滑化周波数帯域制限エネルギーの分散に基づいて音声信号内の音声の開始及び終了点を決定する手段が、

前記平滑化周波数帯域制限エネルギーが所定のエネルギーしきい値レベル

を越えるとき発生するような音声の開始(B)を決定する手段と、かつ

平滑化波数帯域制限エネルギーの前記分散が所定の下部分散しきい値レベルよりも下に降下するとき発生するような音声の終了(E)を決定する手段とを備えていることを特徴とする請求の範囲1の装置。

10. 前記エネルギーしきい値レベル及び前記下部分散しきい値レベルが予め決定されており、かつ前記音声信号の開始(B)が、前記平滑化周波数帯域制限エネルギーが前記エネルギーしきい値レベルを最初に越える前の時間 $z$ 秒にある点として決定されることを特徴とする請求の範囲9の装置。

11.  $z$ が0と100秒との間であることを特徴とする請求の範囲10の装置。

12. 上部及び下部しきい値レベルが予め決定されており、かつ前記音声信号の終了点(E)が、前記分散が前記下部分散しきい値レベルよりも下に降下する前時間の $z$ 秒にある点として決定されることを特徴とする請求の範囲9の装置。

13.  $z$ が0と100秒との間であることを特徴とする請求の範囲12の装置。

14. 前記音声信号の終了点(E)が、平滑化帯域制限エネルギーの分散が前記下部分散しきい値レベルよりも下に降下する前の最後の時間の間、前記平滑化周波数帯域制限エネルギーが前記エネルギーしきい値レベルの下に降下する時間における点として決定されることを特徴とする請求の範囲9の装置。

15. 平滑化周波数帯域制限エネルギーの前記分散及び平滑化周波数帯域制限エネルギーの履歴に基づいて前記音声信号内の音声の開始及び終了点を決定する手段が訓練神経ネットワークを備えていることを特徴とする請求の範囲1の装置。

16. 前記平滑化周波数帯域制限エネルギーが前記エネルギーしきい値を越えた後 $t$ 秒以内に、平滑化周波数帯域制限エネルギーの前記分散が前記上部分散しきい値を越えないならば、音声の前記開始点が受け入れないことを特徴とする請求の範囲9の装置。

17.  $t$  が0と10秒との間であることを特徴とする請求の範囲16の装置。

18. 入力信号内の音声を認識する装置において、音声信号を受け取る手段と、前記信号を有する音声の開始及び終了点を決定する手段と、前記開始点と終了点との間の前記信号内の音声の内容を決定する手段とを有する前記装置とともに、前記音声の開始及び終了点を決定する手段のための改良が、

前記入力信号内の前記平滑化周波数帯域制限エネルギーを表す値を決定する手段と、

前記平滑化周波数帯域制限エネルギーを表す前記値の分散を決定する手段と、

前記平滑化周波数帯域制限エネルギーの分散及び前記平滑化周波数帯域制限エネルギーの前記履歴に基づいて前記音声信号内の音声の開始及び終了点を決定する手段とを備えていることを特徴とする音声の開始及び終了点を決定する手段のための改良。

19. 入力信号  $\times (t)$  における音声の検出のための装置において、

前記入力信号の平滑化周波数帯域制限エネルギーの分散を決定する手段と

前記平滑化周波数帯域制限エネルギーの前記分散及び履歴に基づいて前記信号内の音声の開始及び終了点を決定する音声間隔決定手段とを備えていることを特徴とする入力信号  $\times (t)$  における音声の検出のための装置。

20. 前記平滑化周波数帯域制限エネルギーが前記入力信号をフーリエ変換を通過させることから得られることを特徴とする請求の範囲19の装置。

21. 前記分散が  $m$  秒の連続期間にわたる前記前記平滑化周波数帯域制限エネルギーから決定されることを特徴とする請求の範囲19の装置。

22.  $m$  が0と10秒の間であることを特徴とする請求の範囲21の装置。

23. 平滑化周波数帯域制限エネルギーの前記分散が、 $m$  秒の平滑化周波数帯域制限エネルギーの合計及び前記  $m$  秒の平滑化周波数帯域制限エネルギーの二乗の合計を保持することによって決定され、かつ新しい分散決定に関しては、平滑化周波数帯域制限エネルギーの二乗の合計が最新の平滑化周波数帯域制限エネルギー

一の二乗を加算し、かつ前 $m$ 秒の前記平滑化周波数帯域制限エネルギー値の二乗を減算することによって更新され、及び前記 $m$ 秒の平滑化周波数帯域制限エネルギーの合計が前記最新の平滑化周波数帯域制限エネルギーを加算し、かつ前 $m$ 秒の前記平滑化周波数帯域制限エネルギー値を減算することによって更新されることを特徴とする請求の範囲1の装置。

24. 信号記録装置を含む請求の範囲1の装置において、前記記録装置が、  
前記信号を受け取る手段と、

最も最近 $m$ 秒のその信号を記憶する手段と、

請求の範囲1の装置によって決定された開始及び終了点に対応する前記記憶された信号の一部を選択する手段とを含んでいることを特徴とする請求の範囲1の装置。

25. 信号記録装置を含む請求の範囲1の装置において、前記記録装置が、  
前記信号を受け取る手段と、

最も最近 $m$ 秒のその信号を記憶する手段と、

前記信号を同時に受け取っている間、前 $z$ 秒の前記信号の一部を選択する手段であって、ここで $z$ が請求の範囲1の装置によって決定されることと、  
を含んでいることを特徴とする請求の範囲1の装置。

26.  $z$ が0と100秒との間であることを特徴とする請求の範囲25の装置。

27.  $m$ が0秒又は0秒よりも大きいことを特徴とする請求の範囲25の装置。

28. 前記平滑化周波数帯域制限エネルギーを表す前記値を決定する前記手段が、

前記周波数帯域制限エネルギーを計算する手段と、

前記平滑化周波数帯域制限エネルギーを生成するために平滑化関数を前記周波数帯域制限エネルギーを適用する手段とを含んでいることを特徴とする請求の範囲1の装置。

29. 前記周波数帯域制限エネルギーを平滑化する前記手段が、

前記周波数帯域制限エネルギーを表す最新値の中央値を計算する手段を備えていることを特徴とする請求の範囲28の装置。

30. 前記周波数帯域制限エネルギーを平滑化する前記手段が、

前記周波数帯域制限エネルギーを表す最新値の平均値を計算する手段を備えていることを特徴とする請求の範囲28の装置。

31. 前記周波数帯域制限エネルギーを平滑化する前記手段が、

前記周波数帯域制限エネルギーの速い分散を抑圧するフィルタを適用する手段を備えていることを特徴とする請求の範囲28の装置。

**【発明の詳細な説明】****音声検出装置****技術分野**

本発明は、一般的には、音声セグメント及び非音声雑音又は背景セグメントの両方を含んでいる入力オーディオ信号内の音声を含んでいるセグメントの開始及び終了を検出するための装置に関する。

**従来の技術**

実時間での音声の検出は、音声起動テープレコーダ、返答機、自動音声認識装置、音楽から音声を取り除く処理器（これらに限定されないが）を含む多くの装置に必要な構成要素である。これらのアプリケーションの多くは、分離不可能に音声と混合した雑音を有する。音声の検出は、エネルギーレベルがプリセットしきい値よりいつ上に上がり又はプリセットしきい値よりも下に下がるかを単に検出する従来装置によって提供されるよりもより精巧な音声検出能力を必要とする。

自動音声認識の分野では、音声検出構成要素は最も決定的に重要である。実際には、音声信号の内容を決定するために一般に使用されるパターンマッチングにおけるエラーからよりも多くの音声認識エラーが音声検出におけるエラーから生じる。一つの提案された解決法は、認識装置が常に特定のワードが聞こえないかと耳を常に澄ますワードスポッティング技術を使用することである。しかしながら、ワードスポッティングが音声認識の前に行われなければ、全体のエラー率は高くなり得る。

多くの音声検出装置は、エネルギー、ピッチ及びゼロ交差のような入力のいくつかのパラメータに基づいている。音声検出器の性能は、背景雑音に対するそのパラメータの強さに主に依存する。実時間音声検出に関しては、パラメータは信号から速く抽出されねばならない。

**発明の開示**

本発明の目的の一つは、入力の到達と同じ歩調で進むのに十分高速に、すなわち実時間で動作できる音声の検出のための装置を提供することにある。

本発明の他の目的は、従来のデジタル信号処理回路板で実施され得る音声の



検出のための装置を提供することにある。

本発明の他の目的は、音声と混合されたいろいろな種類の雑音にもかかわらず有効である音声の検出のための装置を提供することにある。

本発明の他の目的は、分離ワード自動音声認識装置、連続音声認識装置（文章の語句間の区切りを検出するための）、音声制御型テープレコーダ、返答機及び背景雑音又は音楽を有する記録に埋め込まれた音声の処理（これらに制限されないが）を含むいろいろなアプリケーションのための音声検出装置を提供することにある。

本発明のこれら及び他の目的は、入力信号における音声を検出する装置を備えることによって達成される。該装置は、前記信号内の平滑化周波数帯域制限エネルギーを表す値を決定する手段と、前記信号の平滑化周波数帯域制限エネルギーを表す値の分散を決定する手段と、前記平滑化周波数帯域制限エネルギーの分散及び前記帯域制限エネルギーの履歴に基づいて前記信号内の音声の開始及び終了点を決定する手段とを含んでいる。

本発明は、入力音声信号内の音声の開始及び終了を検出するために平滑化周波数帯域制限エネルギーの分散及び平滑化周波数帯域制限エネルギーの履歴を利用する。平滑化周波数帯域制限エネルギーの分散は、音楽の背景に対するリード歌手のような難しい背景で発生する前景音声が比較的低い変動の“ノイズフロア”より上のエネルギーレベルの顕著な変動を生じる観察に基づいて使用される。背景のレベルは高いかもしれないけれども、この効果が生じる。分散はそのエネルギーの変動を定量化する。

好ましい実施例によれば、装置は、ハミングウィンドウ及びフーリエ変換を使用して平滑化周波数帯域制限エネルギーを計算する。分散は、シフトレジスタに記憶された平滑化周波数帯域制限エネルギーからの時間の関数として計算される。

音声の開始及び終了点を決定するために、装置は、平滑化周波数帯域制限エネルギーと所定のエネルギーしきい値とを、時間の関数としての分散と2つの所定のしきい値レベル、すなわち上部分散しきい値レベル及び下部分散しきい値レベル

とを比較する。もし平滑化周波数帯域制限エネルギーがエネルギーしきい値を越えるならば、装置は、音声を開始されたことを試験的に決定する。

しかしながら、指定された時間量後に、分散が上部分散しきい値レベルの上にその後上昇しないならば、音声の開始の試験的な決定が廃棄される。平滑化周波数帯域制限エネルギーが、エネルギーしきい値を超過するときと分散が上部分散しきい値を超過するときとの間の時間中、装置は信号を開始（B）音声状態にあるものとして特徴づける。一旦分散が上部しきい値レベルを越えると、装置は信号を音声（S）状態にあるものとして特徴づける。最後に、分散は下部分散しきい値レベルよりも下に降下するとき、音声の終了点が決定される。

他方、平滑化周波数帯域制限エネルギーの最新の履歴及び時間の関数としてのその分散は、訓練神経ネットワークへの入力として使用され、その単一の2進出力は、音声が進んでいるか否かを示している。

分散を試験するために上部及び下部しきい値レベルを使用することによって、音声を検出するときのエラー率は最小にされる。開始点を試験的に決定するために平滑化周波数帯域制限エネルギーのレベルを使用することによって、音声の真の開始と音声検出装置の応動との遅延が最小にされる。音声が存在するかどうかを示すために神経ネットワークを使用することによって、装置は、多くのいろいろな種類の雑音の中で音声を検出できる。

好ましくは、平滑化周波数帯域制限エネルギーの分散及び平滑化周波数帯域制限エネルギーの履歴に基づいて音声の開始及び終了点を決定するための入力信号の処理が実時間で実行され得るように、装置は集積回路ハードウェアの内部で実施される。

#### 図面の簡単な説明

本発明の正確な特質、並びにその目的及び利点は、同じような参照番号がそれの図の至る所の同じような部品を示している添付図面とともに考察されるとき、下記の詳細な説明を参照することによって容易に明かになる。図面において、

図1は、本発明の好ましい実施例による音声検出装置を使用する自動音声認識装置のブロック図を提供する。

図2は、図1の音声検出装置のブロック図である。

図3は、図1の音声検出装置によって使用される平滑化周波数帯域制限エネルギーの分散を決定する方法を示すフローチャートを提供する。

図4は、図2の音声検出装置を示す状態図である。

図5は、典型的な入力信号である。及び

図6は、音声の開始及び終了点を決定するときに神経ネットワークの使用を示す第2の実施例における図2の一つの決定ユニットのブロック図である。

発明を実施するための最良モード

下記の説明は、いかなる当業者も発明を作成し、使用し、自分の発明を実施する発明者によって熟考された最良モードを記載することを可能にするために与えられる。しかしながら、いろいろな変更例は、本発明の一般的な原理が、入力信号の平滑化周波数帯域制限エネルギーの分散に基づいて音声の開始及び終了点を検出する音声検出装置を提供するために特にここに規定されているので、容易に当業者に明かのままでしょう。

本発明を使用する分離ワード自動音声認識システムのためのプロセッサは図1に示される。マイクロホンからのアナログ入力101は、電圧増幅され、サンプリング周波数に等しい速度（一般に、10,000サンプル/秒）でアナログ/デジタル変換器102によってデジタル形式に変換される。結果として生じるデジタル信号103は、6.5536秒の音声—いかなる単一の音声の発声よりも長い期間まで記憶できるメモリ領域104に保管される。104の容量が超過されるならば、古いデータは、新しいデータが保管されるとき抹消される。したがって、104は最も最新の6.5536秒の入力データを含んでいる。デジタル信号103はまた音声検出装置105への入力として役立つ。出力決定信号106は、ゲート107をトリガして音声を含むように

105によって決定されたメモリ104の一部を出力108に送る。異なるアプリケーションに関しては、バッファ104の長さは修正され得て、返答機のようないくつかのアプリケーションでは、バッファ104は排除され得て、信号106はテープ駆動を直接制御できる。他方、バッファ104は単に数ミリ秒の遅延線であってもよい。

音声検出装置105は、図2に詳細に示されている。図1のデジタル入力信号103は、図2の入力信号201として示されている。信号201は、入力のある周波数の連続サンプル（例えば、256）を保持する遅延線に入力する。遅延線が満杯にされると、周波数帯域リミタ203は信号を処理し始める。入力データ201の $n f / 2$ （例えば128）の新しいサンプルが受け取られるとき、遅延線202は128のサンプルを右にシフトし、128の最も古いサンプルを抹消し、左半分を128の新しいサンプルで満たす。このように、シフトレジスタ202は、常に入力の256の連続サンプルを含み、以前の内容と50%重複する。用意するための128の新しいサンプルの時間の単位はフレームであり、1つのフレームは例えば、0.0128秒である。

周波数帯域制限エネルギーは203で計算される。ハミングウィンドウによって遅延線の要素を乗算後、フーリエ変換205は202の内容の周波数スペクトルを抽出する。250Hzと3500Hzとの間の周波数、すなわち最も重要な音声情報を含んでいる帯域、に対応するスペクトル成分は、206でデシベルの単位に変換され、207で一緒に合計され、図2で信号251として示される周波数帯域制限エネルギーを発生する。

他方、周波数帯域制限エネルギーは周波数スペクトル変換器の一部を合計する以外の方法によって計算され得る。例えば、入力信号は、畳み込みによって又は再帰型フィルタを通過させることによってデジタル的にフィルタリングされ、そのエネルギーは下記の方法によって測定され得る。これは、図2の202及び203の全部を取り替える。

同様に、帯域制限は、アナログフィルタから直接得られるエネルギーを有するアナログ領域で又は下記の方法によって実行され得る。アナログ帯域リミタは、バンドパスフィルタ、ローパスフィルタ、又は他のスペクトル整形フィルタから成ってもよく、又は増幅器あるいはマイクロホンに固有の周波数制限から生じてもよく、もしくは偽信号防止フィルタの形を取ってもよい。エネルギーは直接フ

ィルタから又は下記のパラグラフに記載された方法によって得られてもよい。これらの代替の技術のいずれから結果として生じる信号は、以下、周波数帯域制限信号と称される。

周波数帯域制限エネルギーのエネルギーとともに一般に単調に分散する量は、以下、周波数帯域制限エネルギーと呼ばれる。図2に記載された方法の代わりに、周波数帯域制限エネルギーは、(a) 短い時間間隔にわたって周波数帯域制限信号の分散を計算すること、(b) 短い時間間隔にわたる周波数帯域制限信号の絶対値、大きさ、整流値、又は他の等しい電力の二乗を合計すること、あるいは(c) 短い時間間隔にわたる周波数帯域制限信号の値のピーク、大きさ、整流値、又は他の等しい電力の二乗を決定することによって計算されてもよい。

本発明の好ましい実施例に関して続けると、周波数帯域制限エネルギーは平滑化モジュール220によって平滑化される。周波数帯域制限エネルギーは、最初に遅延線259に入る。フレーム毎に、この例では12.8ミリ秒で、この遅延線は新しいサンプルを受け取り、残りのサンプルを右に1だけシフトする。この例では、その長さは0.128秒に対応する10フレームである。より短い長さは音声検出装置の応答時間を減少する。より長い長さは、装置を衝撃雑音に対してより強くする。

平滑化計算ユニット250は遅延線259の内容の平均値を計算し、その値は平滑化周波数帯域制限エネルギー208である。

他方、平滑化計算250は、遅延線259における値の中央値を計算することによって、又は遅延線259の内容の短い衝撃分散を平滑化あるいは違った方法で抑制するための効果を有するいかなる関数をも計算することによって実行されてもよい。低下した場合には、遅延線259の長さは1であり得て、信号251は出力208に直接送られ得るので、平滑化周波数帯域制限エネルギー208は周波数帯域制限エネルギー251と同一である。

平滑化周波数帯域制限エネルギーは遅延線209に入る。平滑化計算250は遅延線259の内容の急速な変化を取り除くための効果を有しているので、分散計算のための遅延線209は、フレーム毎に1回よりも遅い速度で新しい値を受け取ってもよい。遅延線209は、各新しい入力が入力されると、1だけ右にシフトする。より長い遅延線は、音声を終了させることを宣言する前に発声内部により長い区切りを可能

にする。より短い遅延線は音声の終了に対する音声検出器の応答を速くする。この遅延線の長さは  $nv$  である。これは、この例では0.51秒の区切り長さに対応する40であり、下記のようになる。

$$nv = \frac{(\text{区切り長さ}) \times (\text{サンプリング周波数})}{(nf/2)}$$

分散計算ユニット210は遅延線209における値の分散を計算する。平滑化周波数帯域制限エネルギーの分散  $V$  は、下記のようになる。

$$V = g(A, B)$$

ここで、

$$g(A, B) = \frac{A}{nv} - \frac{B \times B}{nv \times nv}$$

及び

$$\begin{aligned} f &= nv \\ A &= \sum_{f=1}^{f=nv} (BLE(f) \times BLE(f)) \\ f &= 1 \end{aligned}$$

及び

$$\begin{aligned} f &= nv \\ B &= \sum_{f=1}^{f=nv} BLE(f) \\ f &= 1 \end{aligned}$$

及び

$V$  は分散計算210の出力211である。

及び

$BLE(f)$  は、位置  $f = nv, \dots, 3, 2, 1$  での遅延線209の内容である。

$BLE(1)$  は、最も古い  $BLE$  値であり、かつ  $BLE$  は平滑化周波数帯域制限エネルギーである。

及び

分散211及び平滑化フィルタリング帯域制限エネルギー208は、その動作が図4

及び図5に示されている決定ユニット212を駆動する。

図3は、分散Vを計算するためのより速い方法を示し、分散計算210及び遅延線209を取り替える。このより速い技術は、再計算するよりも下記のように量A及びBを更新する。

$$A' = A + [BLE(nv) \times BLE(nv)] - [BLE(0) \times BLE(0)]$$

$$B' = B + BLE(nv) - BLE(0)$$

ここで、

A' は302として示されるAに対する更新値である。

B' は303として示されるBに対する更新値である。

及び

BLE(nv)は図2の208からの最新の平滑化周波数帯域制限エネルギー301であり、かつBLE(0)は最古の平滑化周波数帯域制限エネルギー304である。

BLEの二乗は遅延線305で遅延される。この遅延線は、304からの値を二乗することによって取り除かれ、取り替えられ得る。遅延線305及び306は、初期化の際ゼロにクリアされるべきである。同様に、遅延線306及び305が図2の遅延線209よりも一つ長いことに注目せよ。

図6は、神経ネットワークを使用する決定ユニット（図2の212）のブロック図である。神経ネットワーク620への入力は、その1.28秒の音声からの周波数帯域制限エネルギーのいくつかのサンプル及び平滑化周波数帯域制限エネルギーの分散である。遅延線603は、前1秒の平滑化周波数帯域制限エネルギー602を蓄積し、レジスタ604は周波数帯域制限エネルギー601の分散を蓄積する。神経ネットワークの出力621は、現在のフレームが音声を含んでいるか否かを示す2進決定である。これは図2の214に対応する。

他方、決定ユニットはしきい値処理方式を使用できる。図4は、音声の存在を検出するために分散（図2の211）及びエネルギー（図2の213）を使用する決定ユニットのための状態図を示している。図5は、平滑化周波数帯域制限エネルギーSBLE及び音声信号の平滑化周波数帯域制限エネルギーの分散VSBLE並びに状態図を理解するときに補助として対応する状態の例を示している。各フレーム、こ

の例では0.0128秒でこの状態図の遷移が行われる。

状態図は、N—すなわち雑音—状態（502）で始まる。SBLEがエネルギーしきい

値510よりも下にある限り、遷移402が行われ、状態Nは抜けられない。SBLEがエネルギーしきい値510より上に上昇するとき、遷移403が行われ、状態B（音声の試験的な開始503）が達せられる。このように、エネルギーは装置を速くトリガするために使用される。状態Bが達せられるとき、装置は、音声は2、3ミリ秒前に開始したことを決定する。この時間量 $z$ は一般に遅延線259の長さに等しい。

プリセット時間量に関しては、状態Bは抜けられない。すなわち、遷移404が行われる。もしこの時間があまり短いならば、開始点推測が遅れすぎ、音声の頭が切断される。この時間がより長くなると、音声の開始に対する音声検出器の応答は、不正確でないけれども、遅延されるようになる。もしそれが遅延線209の長さよりも長いならば、装置は音声を完全に見落とすかもしれない。この例では、時間は175ミリ秒である。この時間の終わりに、VSBLEは、それが506、すなわち上部分散しきい値を越えたかどうかを調べるために検査され、状態Bが抜けられる。もしVSBLEが上部分散しきい値より下にあるならば、遷移406が行われ、試験的な開始点が廃棄され、装置はN状態に戻る。もしVSBLEが上部分散しきい値より上にあるならば、遷移405が行われ、装置は、音声は装置にこれまで入力されてきており、現在入力されていることを決定してきていることを意味するS状態504に入る。

VSBLEは下部分散しきい値501より上に留まる限り、遷移407が行われ、状態Sが抜けられない。VSBLEが下部分散しきい値よりも下に落ちるとき、遷移408は装置を音声の終了が検出されたことの信号を送出するE状態にする。SBLEが、E状態が達せられる前に最後の時間の間エネルギーしきい値よりも下に落ちる点にあるように、音声の終了が決定される。次のフレームで、装置はN状態に戻る。

図1のゲート107の後の装置が自動音声認識装置であるならば、現在の状態を図2の線214上で送り、ゲート107を制御するためにそれを図1の106に接続する



ことによって、自動音声認識装置は実時間で入力信号を処理できる。唯一の遅延は、開始点を決定するために音声検出器によってかかる時間である。音声の状態Bで自動音声認識装置に送られ得るならば、すなわち、ゲート又は認識装置が遷移406が行われる場合に入力音声を取り消す能力を有するならば、自動音声認識装置は、遅延線259の長さにおよそ等しい遅延を有する音声进行处理し始めることができる。

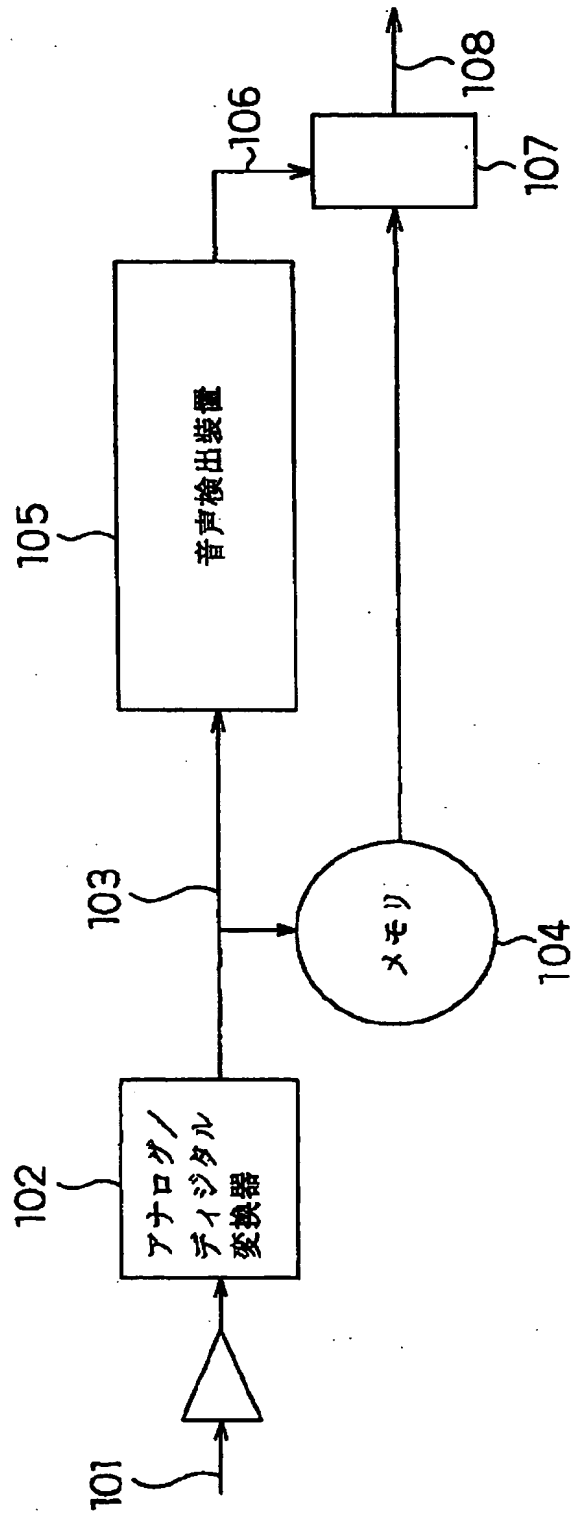
記載されているものは入力信号内の音声の存在を検出する装置である。装置は

信号内の平滑化周波数帯域制限エネルギーの分散に基づいて音声の開始及び終了点を計算する。平滑化周波数帯域制限エネルギーの分散を利用することによって、音声の存在が実時間で有効に検出される。装置は、セグメントが抽出され、さらに処理される得るように、音声を含んでいる記録のセグメントを検出するのに特に有用である。

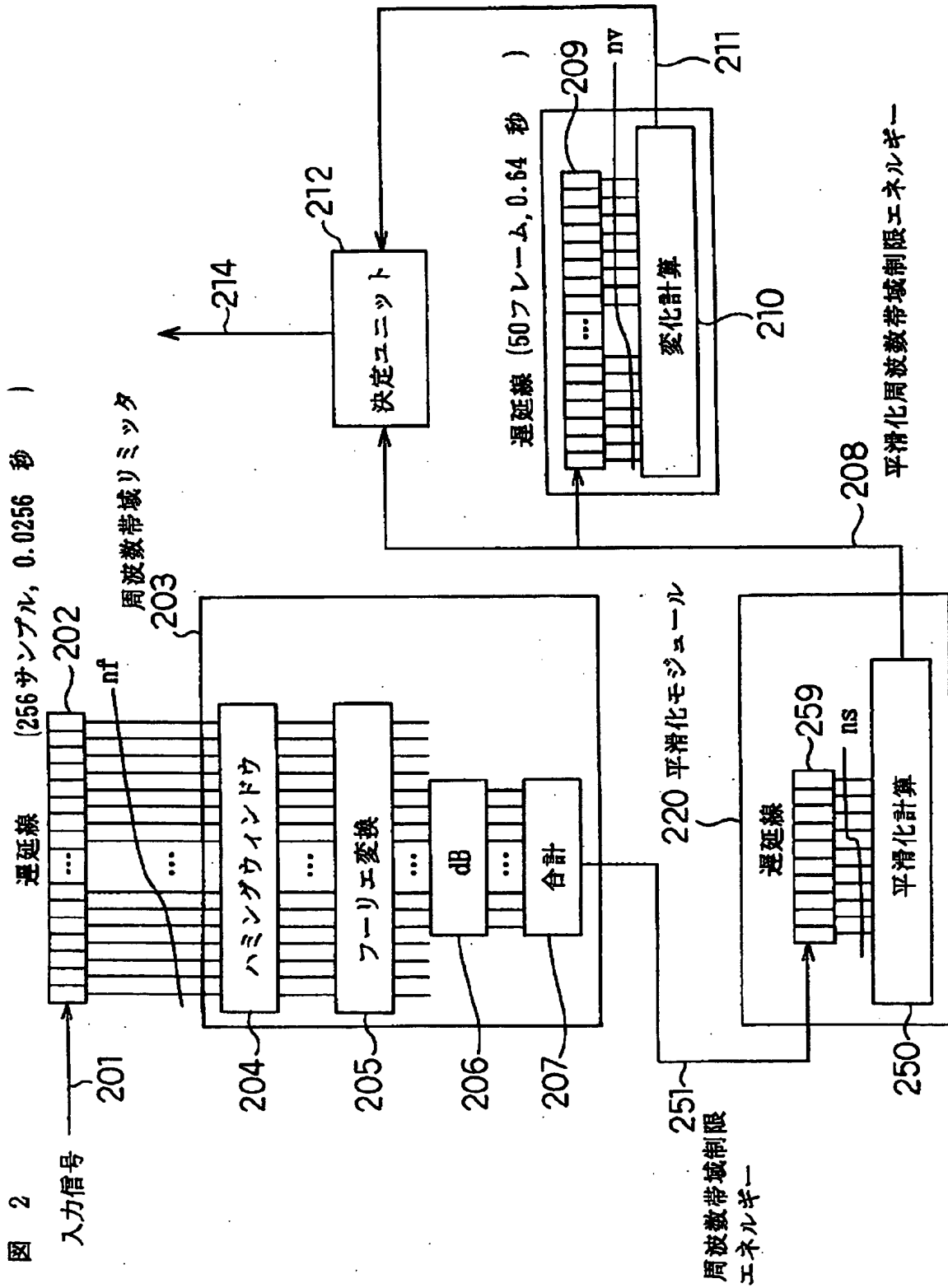
当業者は、前記の好ましい実施例のいろいろな変更及び修正が本発明の範囲及び精神を逸脱しないで構成され得ることを理解するだろう。したがって、添付クレームの範囲内で本発明は特にここで記載されたもの以外に実施され得ることが理解されるべきである。

【図1】

図 1

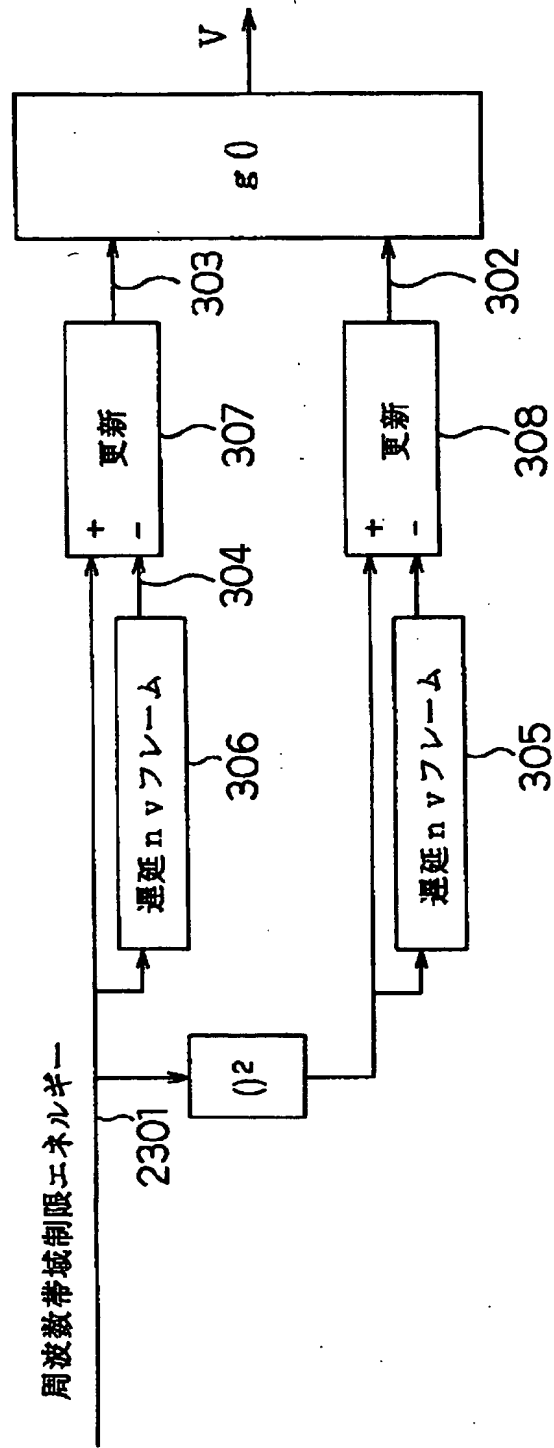


【図2】



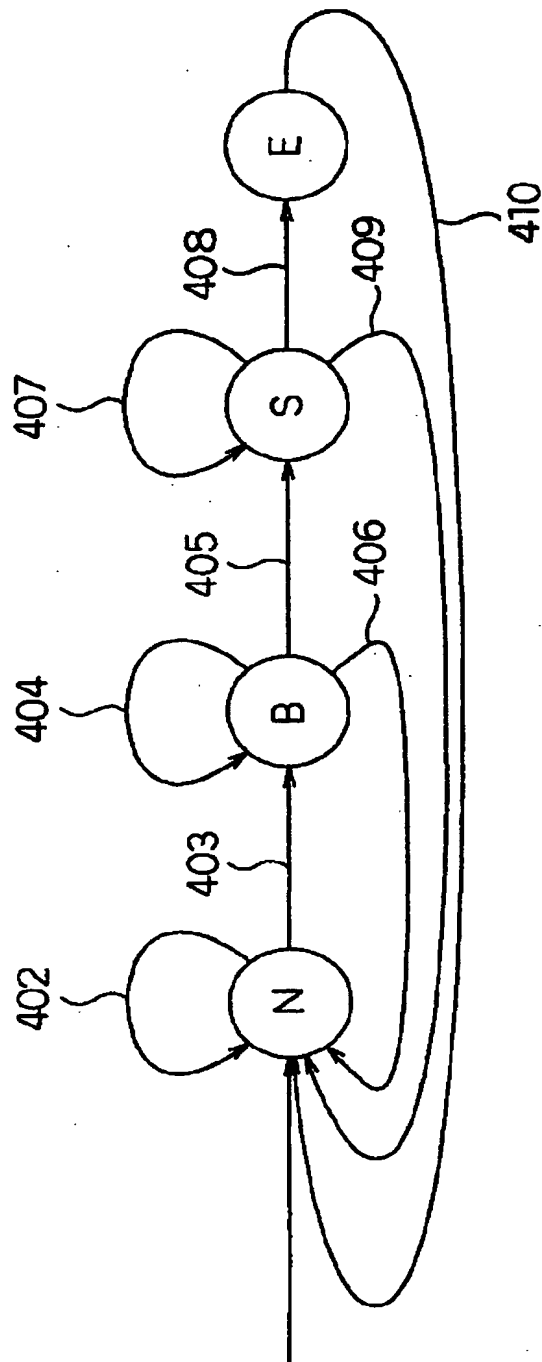
【図3】

図 3



【図4】

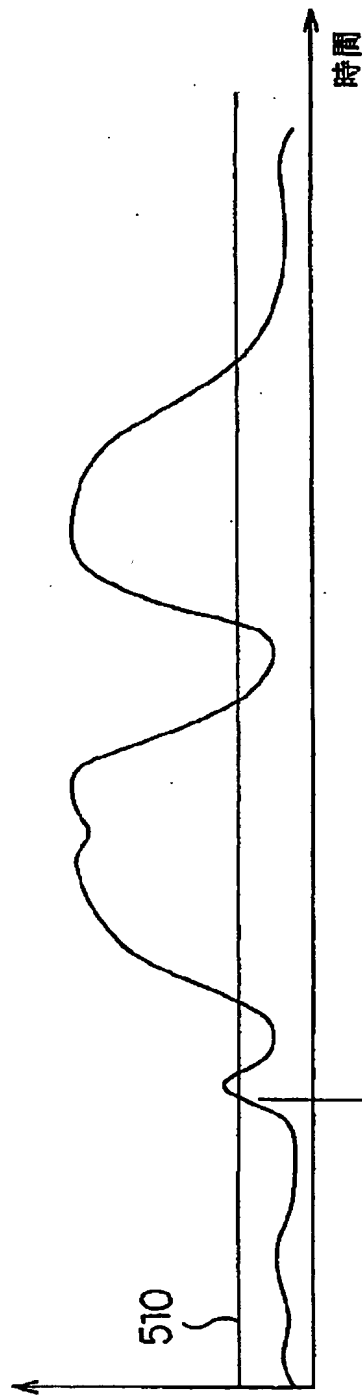
図 4



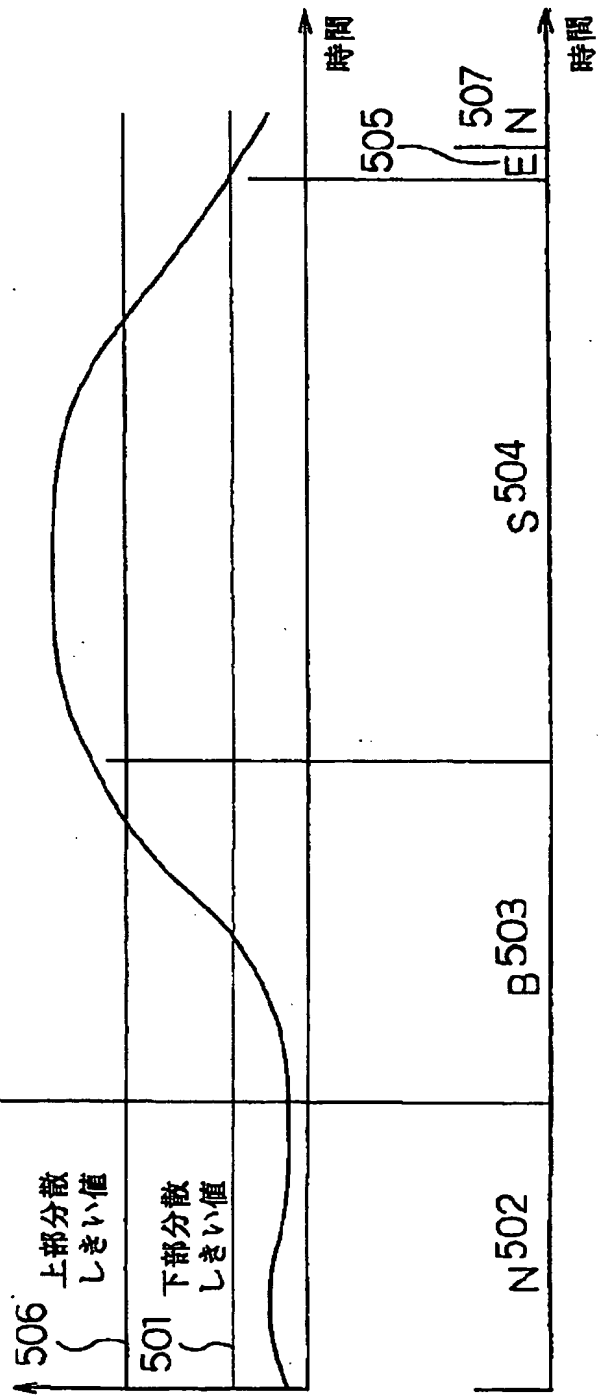
【図5】

図 5

平滑化  
帯域制限  
エネルギー  
(SBLE)



分散平滑化  
帯域制限  
エネルギー  
(VSBLE)



状態

【図6】

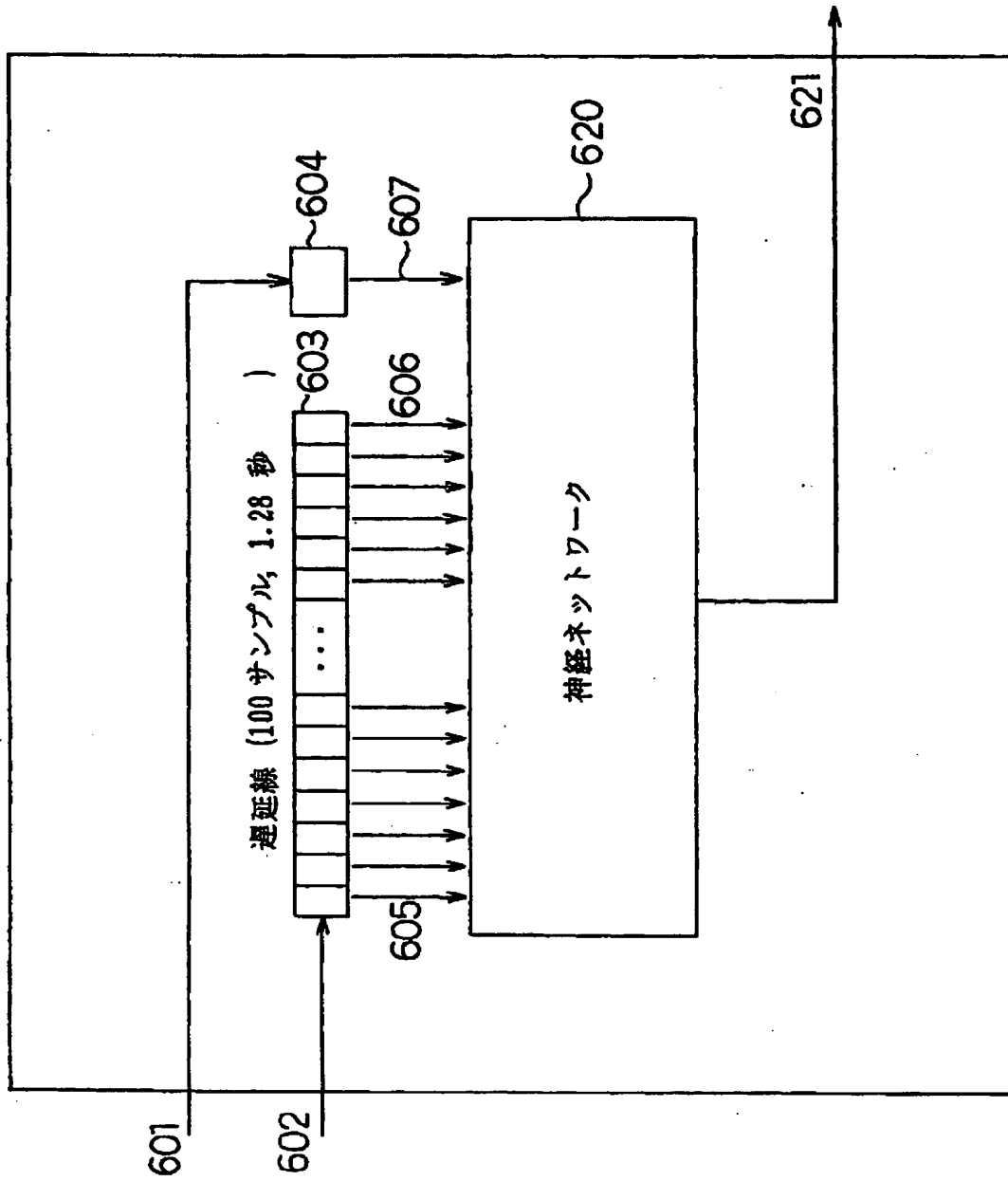


図 6

【国際調査報告】

## INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER IPC 6 G10L3/00		International Application No PCT/JP 94/01181
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC 6 G10L		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US,A,4 441 203 (M.C. FLEMING) 3 April 1984 see abstract see claims 1-5	1, 18
A	EP,A,0 138 071 (SIEMENS) 24 April 1985 see abstract see claims 1-5	1, 18
A	EP,A,0 167 364 (AT&T) 8 January 1986 see abstract see claims 1-9	1, 18
A	EP,A,0 111 947 (PHILIPS) 27 June 1984 see abstract	3
<input type="checkbox"/> Further documents are listed in the continuation of box C. <input checked="" type="checkbox"/> Patent family members are listed in annex.		
* Special categories of cited documents : "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "U" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. "&" document member of the same patent family		
Date of the actual completion of the international search 29 March 1995		Date of mailing of the international search report 05.04.95
Name and mailing address of the ISA European Patent Office, P.B. 5811 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016		Authorized officer Daman, M



## INTERNATIONAL SEARCH REPORT

Information on patent family members

Int. Patent Application No

PCT/JP 94/01181

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US-A-4441203	03-04-84	NONE	
EP-A-0138071	24-04-85	DE-A- 3335343 DE-A- 3468765	11-04-85 18-02-88
EP-A-0167364	08-01-86	JP-A- 61032900	15-02-86
EP-A-0111947	27-06-84	DE-A- 3243232 AU-B- 561287 AU-A- 2154683 CA-A- 1206620 JP-A- 59105696 US-A- 4682361	24-05-84 07-05-87 31-05-84 24-06-86 19-06-84 21-07-87